УДК 611.08(611.24)
DOI https://doi.org/10.32782/2663-5941/2024.5.1/35

Панаскін Д.В.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ЗАСТОСУВАННЯ МЕТОДІВ АУГМЕНТАЦІЇ CVAE ДЛЯ ПОКРАЩЕННЯ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ В АНАЛІЗІ ЛЕГЕНЕВИХ ЗВУКІВ

Стаття присвячена дослідженню нейронних мереж в аналізі легеневих звуків, як додаткового захисту для запобігання та розвитку хвороб дихальних шляхів, а також застосуванню методів аугментації CVAE для покращення навчання нейронних мереж. У статті оцінено можливість використання акустики, яка аналізує шум для визначення ознак легеневих патологій. Показано, що для виявлення патологій доцільно використовувати різницю нормованих рівнів шуму та спектри в логарифмічній шкалі для лівої та правої легені. Доведено, що диференціальна діагностика шумів може бути заснована на використанні модифікованої крос-кореляції функції спектрів у логарифмічному масштабі, а також їх миттєві характеристиці за допомогою нейронних мереж. Для збереження різноманіття згенерованих зразків, ми використовували переваги методів аугментації CVAE. Для відображення легеневих звуків ми використовували мережу кодувальника з генерацією реального зображення до прихованого вектора. Потім генератор давав необхідні для реконструкції необроблені пікселі у відповідності характеристикам вихідних зображень із заданим прихованим вектором. При використанні цього методу аугментації можна збільшити різноманітність навчальних даних, що підвищить ефективність.

Представленотехнологіюдвоетапногонавчаннянейронних мереж(CVAE),якевідрізняєтьсявикористанням аугментації зображень для попереднього етапу та точності налаштування вагових коефіцієнтів на основі вихідного набору зображень. На першому етапі навчання здійснюється на аугментованих даних, на другому етапі виконується точне налаштування, що сприяє підвищенню ефективності реідентифікації. Отримані результати можуть слугувати основою для створення багатоканальної автоматизованої системи аналізу акустичного шуму для диференціальної діагностики стану легенів, що дозволить створити сучасну систему моніторингу захворювань органів дихання населення.

Ключові слова: *медицина, штучний інтелект, захворювання легень, аугментація, діагностика.*

Постановка проблеми. В даний час цифрова обробка зображень стала важливою частиною медичної діагностики та має широкий спектр застосування. Аналіз легеневих звуків є важливим кроком у медичній діагностиці. Штучні нейронні мережі, особливо загальні нейронні мережі (CNN), сьогодні є основним інструментом вирішення цієї задачі. Основною проблемою застосування нейронних мереж до вирішення задачі сегментації є необхідність формування порівняно великого набору зображень.

Відомо, що значною мірою на ефективність роботи CNN впливає кількість зображень у навчальній вибірці та їх різноманітність. Недолік тренувальних даних може спровокувати перенавчання CNN, запам'ятовування вихідних даних та нездатність до узагальнення ознак загалом. Вирішенням цієї проблеми може бути концепція перенесення навчання (transfer learning). У такому разі CNN навчається на наборі даних досить великого розміру, наприклад ImageNet (3,2 млн анотованих зображень) [1] або LUPerson (близько 4 млн зображень) [2]. Таким чином, CNN навчається у два етапи. На першому етапі виконується попереднє навчання на великому наборі даних, а на другому етапі коригуються на навчальній вибірці для вирішення конкретної задачі, наприклад виявлення та класифікації об'єктів різних класів [3], застосування методів аугментації.

Дослідження показали, що дворазове збільшення отриманої навчальної вибірки шляхом повторної подачі зображень на вхід нейронної мережі (з виконанням аугментації) дозволяє поліпшити якість сегментації (у середньому на 1–2% за метрикою IoU) і не перенавчити модель. Отже, підготовлений набір даних ще не є оптимальним. Одним із способів збільшення обсягу корисної інформації є застосування різноманітних перетворень вихідних даних. Якщо вихідними даними є зображення, то такими перевагами є афінні перетворення, зміна яскравості, контрастності і т.д., що входять в набір методів аугментації.

Метод регуляризації з використанням аугментації CutMix заснований на застосуванні пакета зображень, яке передбачає заміщення області фрагмента одного зображення областю такого ж розміру іншого зображення з цього пакета [4] (рис. 1, b). Застосування CutMix дозволяє CNN підвищити стійкість до перекриття і при цьому зберігати ту ж частину інформації, що втрачається при використанні методу random erasing. В алгоритмі Mixed Single Thumbnail (MST) [5] фрагмент зображення заміщається на зменшену копію фіксованого розміру іншого зображення з пакета (рис. 1, c). Це дозволяє тимчасово вивчати два зображення при вилученні ознак, підвищуючи стійкість до оклюзій і низької роздільної здатності зображень. Одночасно два зображення з пакету обробляються і при використанні аугментації mixup [6], проте їх об'єднання виконується за допомогою лінійної інтерполяції (рис. 1, d). Такий підхід дозволяє покращити узагальнюючу здатність навченої CNN і знижує чутливість до запам'ятовування помилкових міток, які можуть зустрічатися у існуючих наборах даних.

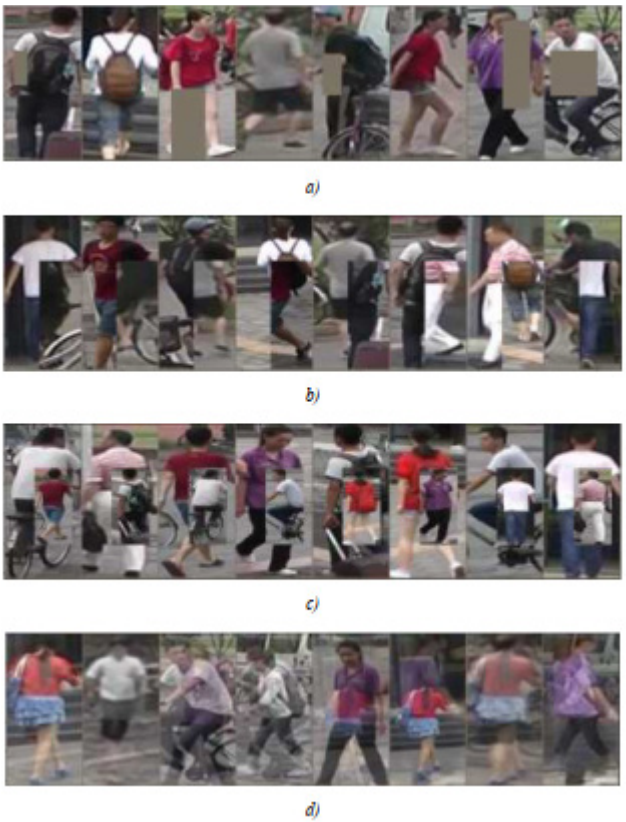

Рис. 1. Приклади використання методів аугментації даних: а) довільне стирання; б) CutMix; в) MCT; г) плутанина [6]

Generative Adversarial Network (GAN) [7] є популярною генеративною моделлю. Вона одночасно тренує т-wo моделі: генеративна модель для синтезу зразків і дискримінаційна модель для розрізнення природних і синтезованих зразків. Однак модель GAN складні на етапі навчання та генеруються зразки часто далекі від природних. Наприклад, Wasserstein GAN (WGAN) [8], використовує Earth Mover Distance як ціль для навчання GAN. Популярності набувають методи засновані на CVAE, які включають генерацію умовного обличчя, At-tribute2Image [9], синтез тексту в зображення, прогноз статичних зображень, а також умовний синтез зображень [10]. Усі вони досягають вражаючих результатів.

Аналіз останніх досліджень і публікацій. Результати досліджень, подані в роботах [11-12], показують, що під час навчання задачі класифікації нейронна мережа більше уваги приділяє текстурі об'єкта, а не його формі. Так, в [11] зазначено, що якщо образ кота заповнити текстурою шкіри слона, то CHP розпізнає клас виявленого об'єкта як «слон», тоді як людина вважатиме, що це кіт, т. е. в людини вирішальне значення має саме форма об'єкта, а не його текстура. У роботі [13] пропонується метод аугментації, що складається з кількох етапів. Спочатку CHP навчається на вихідних даних до того моменту, коли значення функції втрат перестають зменшуватися. Після навчання на зображеннях визначаються області, які являються найбільш важливими для ухвалення рішення нейронною мережею.

У роботі [14] для аугментації пропонується метод random erasing, у якому зображення вибираються для перетворень з вихідного набору даних із застосуванням генератора псевдовипадкових чисел (ГПВЧ). На основі ГПВЧ також визначаються розмір та координати фрагмента зображення, пікселі якого заповнюються нульовими чи випадковими значеннями. Тому при завантаженні пакета зображень на різних етапах навчання одне й те саме зображення може бути представлене як у вихідному вигляді, так і з різними-зміненими фрагментами, що дозволяє підвищити стійкість CNN до оклюзій, проте при цьому частина інформації втрачається.

Ефективність попередніх моделей CNN в медичних цілях підтверджена результатами досліджень, опублікованими у роботах [15–16], і зумовлена тим, що на етапі попереднього навчання виділяються ознаки, що несуть основну інформацію про зображення: вертикальні та горизонтальні лінії, колір, текстура та форма об'єктів та ін. Навчання нейронних мереж широко поширене не тільки для вирішення завдань класифікації та розпізнавання

об'єктів на зображеннях, але і для сегментації зображень, пошуку ключових точок та обробки тексту. Застосування таких мереж обмежується наявністю попередньо навчених моделей для кінцевого числа архітектур CNN та високим споживанням обчислювальних ресурсів та часу.

Іншим вирішенням проблеми перенавчання є регуляризація. У машинному навчанні під регуляризацією розуміють додавання обмежень до архітектури нейронної мережі, або до навчальних наборів даних [17]. Прикладами регуляризації є: проріджування нейронних зв'язків CNN; використання різних функцій активації; L1- та L2-регуляризації; аугментація даних, коли частина зображення може бути видалена або замінена іншою інформацією. Аугментація застосовується для збільшення навчальної вибірки на основі наявних даних [18] за рахунок перетворення зображення. При цьому можуть використовуватися такі перетворення, як зміна яскравості та контраст, дзеркальне відображення, поворот, розмиття та ін. Кількість зображень при аугментації не збільшується, а різноманітність досягається за рахунок того, що на різних етапах навчання до зображень застосовуються різні перетворення. Розширення навчальної вибірки дозволяє поліпшити узагальнюючу здатність CNN і збільшити точність роботи, у тому числі при різних факторах, таких як висока варіація освітлення, низька роздільна здатність, перекриття об'єктів та ін.

Постановка завдання. Мета статті – дослідження нейронних мереж в аналізі легеневих звуків, як додаткового захисту для запобігання та розвитку хвороб дихальних шляхів, а також застосування методів аугментації CVAE для покращення навчання нейронних мереж.

Матеріали та методи. Статистичне опис акустичних шумів, що виникають у процесі дихання, описаний на використанні вкладених двокомпонентних випадкових процесів $\{\vec{S}(t),\theta(t)\}$, у яких одна компонента $\vec{S}(t)$ безперервна, а інша $\theta(t)$ дискретна. Ці компоненти є залежними та, у загальному випадку, не Марківськими. Раніше подібний підхід використовувався для опису нестаціонарних негаусових перешкод, що створюються відбиттями радіохвиль від поверхні моря, суші, «ясного неба». Він виявився продуктивним і для опису радіолокаційного відбиття від малорозмірних надводних цілей. Аналогічний підхід використовується і для опису акустичних шумів, викликаних вітром, дощем, листям дерев, кроками людей та тварин, а також звуками пострілів. У процесі дихання виділяються дві фази: вдиху та видиху. Зміна фазових станів процесу описується квазі детермінованою функцією, на яку не накладається жорсткі обмеження на розподіл часів існування процесу у кожному з фазових станів. Для опису процесу всередині фазового стану використовували стандартні моделі гаусових процесів.

Виклад основного матеріалу. Технології виділення сигнатур шумів за патології легень. *а) Використання усереднених спектрів*. Класичний спектральний аналіз як повної структури спектру шумів дихання, так і окремих його фаз (вдиху та видиху) дозволяє здійснювати диференціальну діагностику патології. На рис. 2а наведено спектри шуму везикулярного дихання, ослабленого везикулярного дихання та при початковій стадії пневмонії, а на рис. 2б – різницеві (диференціальні) спектри. Диференціальні спектри дозволяють виявити відмінності в шумах лівої та правої легень, а також пов'язану із цим патологію.

б) Моментні властивості спектрів. Для діагностики можна використовувати моментні характеристики спектрів – його середнє та середньоквадратичне значення частоти, отримані за окремими часовими сегментами:

$$F(t) = \frac{\int_{-\infty}^{\infty} FS(t,F)\,dF}{\int_{-\infty}^{\infty} S(t,F)\,dF} = \int_{-\infty}^{\infty} Fs(t,F)dF;\ \sigma(F) = \sqrt{\left(\int_{-\infty}^{\infty} F^2 s(t,F)dF - F(t)^2\right)}, \quad (1)$$

де $s(t,F) = \frac{S(t,F)}{\int_{-\infty}^{\infty} S(t,F)\,dF}$ – нормований поточний спектр шумів.

Ці характеристики дозволяють оцінювати зміни середньої частоти та ширину спектра шуму при різних фазах дихання. Ще більш інформативним є усереднені спектри і диференціальні спектри різних фаз дихання. Для їх отримання виділяють часову реалізацію акустичних шумів процесу дихання фази вдихів і видихів і для кожної з фаз обчислюються усереднені спектри.

На рис. 3 наведено спектрограми шумів при нормальному диханні та при початковій стадії пневмонії. БПФ оцінювався за сегментами тривалістю близько 0,1С. Патології легень проявляються у часових реалізаціях шуму, їх спектрограмах і спектрах, як поточних, так і середніх, а також моментних характеристиках спектрів, з яких найбільш зручними для аналізу патологій є зміни середніх і середньоквадратичних значень частот. Пропонований підхід відкриває нові можливості для диференціальної діагностики патології легень.

Технологія двоетапного навчання із аугментацією даних. Пропонується використовувати підхід, що включає технологію двоетапного навчання CNN та новий метод аугментації даних. При цьому на першому етапі виконується попереднє навчання на аугментованих даних, а на дру-

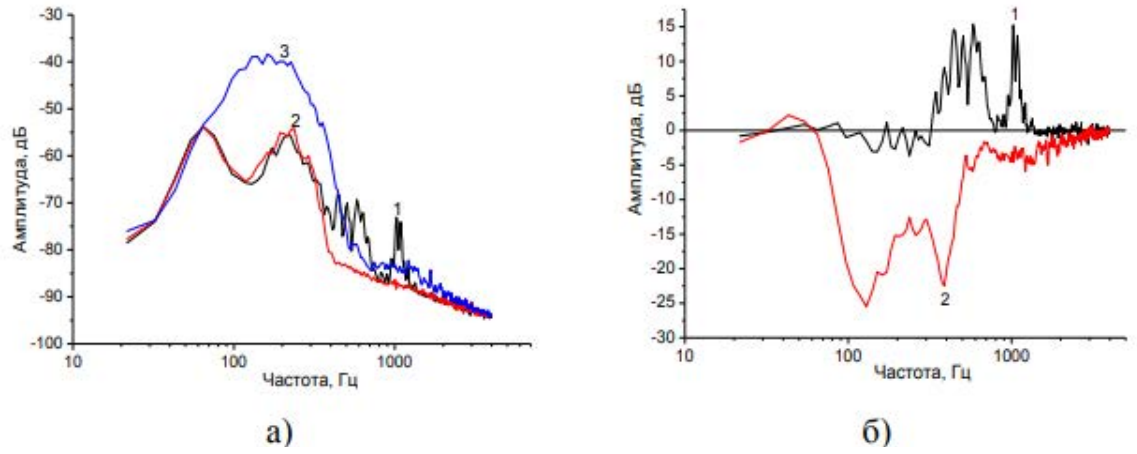

Рис. 2. Спектри (а) при ослабленому (1, 2) везикулярному (3) диханні, а також диференціальні спектри (б) при ослабленому (1) везикулярному (2)

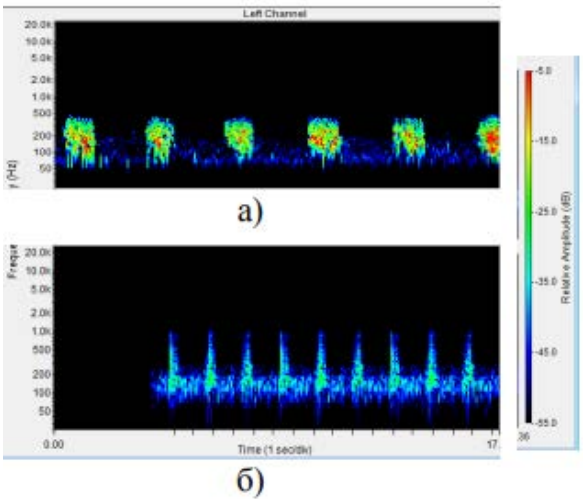

Рис. 3. Спектрограми везикулярного дихання (а) та при пневмонії (б)

гому етапі застосовується точне налаштування CNN, при якій вагові коефіцієнти коригуються на вихідних зображеннях із меншою швидкістю навчання. Це дозволяє отримувати найбільш надійні відмітні ознаки. Для збільшення різноманітності навчальної вибірки на основі наявних даних пропонується використовувати циклічний зсув пікселів по вертикалі та горизонталі, виключення кольоровості та заміщення фрагмента іншим зменшеним зображенням.

Циклічний зсув та виключення кольоровості застосовуються до окремих зображень, а заміщення фрагмента здійснюється зменшеною копією іншого пакета, що подається на вхід CNN. Попереднє навчання моделі та точне налаштування. Для зниження значення функції втрат під час навчання, високі рівні якої викликані неправдоподібністю даних щодо тестових даних, та підвищення точності повторної ідентифікації пропонується застосовувати двоетапне навчання. При цьому використовували наступні особливості: швидкість тренування класифікаційного шару вища за швидкість всіх інших шарів CNN, на першому етапі здійснюється попереднє навчання із застосуванням аугментації даних, на другому етапі тренування CNN триває тільки на вихідному наборі даних.

Тому ми запропонували використовувати CVAE для навчання нейронних мереж. Як показано на рис. 4, запропонований нами метод складається з чотирьох частин: 1) мережа кодера E; 2) родитивна мережа G; 3) дискримінаційна мережа D; і 4) мережа класифікації C. Функція мереж E і G така ж, як і в умовно-варіаційному автокодері (CVAE) [19]. Ан-мережа кодера E відображає вибірку даних x на латентне представлення передачі z через вивчений розподіл P (z|x, c), де c – категорія даних. Твірна мережа G генерує зображення x′ шляхом вибірки з вивченого розподілу P (x|z, c).

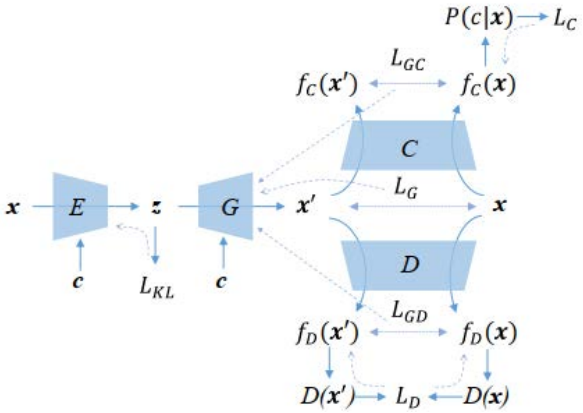

Рис. 4. Ілюстрація структури нашої мережі. Наша модель містить чотири частини: 1) мережа кодера E; 2) твірна мережа G; 3) класифікаційна мережа C; та 4) дискримінаційна мережа робота D

Функція мережі G і D така ж, як в генеративній змагальній мережі (GAN). Мережа G намагається дізнатися реальний розподіл даних за допомогою дієнтів, заданих дискримінаційною мережею D, яка навчається розрізняти «справжні» та «фальшиві» аугментаційні дані. Функціональною мережею C вимірювали зсуви везикулярного дихання P (c|x). Однак комбінація VAE і GAN є недопустимою. Остання робота [20] показує, що навчання GAN страждатиме від проблеми зникнення градієнта або нестабільності мережі G. Тому ми лише пропонуємо використовувати генеративну мережу CVAE для покращення аналізу легеневих захворювань.

Лістинг. Алгоритм вибору зображень та спектрограм везикулярного дихання та дихання при ушкодженнях легень, для яких виконується перетворення.

1. Input: I_{in} – вхідне зображення
2. p – відсоток зображень, до яких застосовується перетворення
3. Transform – перетворення
4. Output: Iout – вихідне зображення
5. Initialization: r ← Rand(0,1);
6. t=p/100
7. if r > t then
8. Iout ← Iin;
9. return Iout.
10. else
11. I_{out} ← Transform(Iin);
12. return I_{out}.
13. end

Для виявлення патологій доцільно використовувати різницю нормованих рівнів спектрів шуму в логарифмічному масштабі для лівого і правого легенів. Для бінарної системи легенів необхідно сформувати координати обмежувальної рамки вирівнювального фрагмента E = (r_x, r_y, r_w, r_h), де r_x, r_y – координати лівого нижнього кута; r_w і r_h – висота та ширина віддаленої ділянки, які визначаються за допомогою CVAE таким чином, що r_h становить (0,25 ... 0,5 H), де H – Висота вихідного зображення. Ширина визначається як: $r_w = \left[\frac{r_h}{\eta}\right]$, де $\eta = \left[\frac{H}{W}\right]$ – співвідношення сторін вихідного зображення. Застосування технології двоетапного навчання та запропонованого методу аугментації надано для CHP ResNet-50, навченої на різних наборах даних, дозволило підвищити точність на 4,18–21,55%.

Висновки. Захворювання легень займають одне з перших місць за втратами працездатності у всіх країнах світу. Дослідження в даній галузі та роботи у цьому напрямі еволюціонують у міру розвитку обчислювальної техніки, типів датчиків, методів штучного інтелекту в галузі діагностики та прийняття рішень, засобів телемедицини. Зображення спектрограм шумів при нормі та патології можна використовувати для виявлення патологій з використанням підході використовуваних при розпізнаванні образів. При цьому обчислення віконного перетворення Фур'є доцільно проводити за сегментами тривалістю близько 0,1 сек. Розглянуто можливість використання різних технологій аналізу акустичних шумів для визначення сигнатур патології легень. Це може стати основою створення багатоканальної автоматизованої системи діагностики стану легень.

Список літератури:

1. ImageNet: A large-scale hierarchical image database / J. Deng [et al.] // 2009 IEEE Conf. on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009. Miami, 2009. P. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

2. Unsupervised pre-training for person re-identification / D. Fu [et al.] // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. Nashville, 2021. P. 14745–14754. https://doi.org/10.1109/CVPR46437.2021.01451

3. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking / E. Ristani [et al.]. 2016. Mode of access: https://doi.org/10.1007/978-3-319-48881-3_2. Date of access: 11.08.2024.

4. CutMix: Regularization strategy to train strong classifiers with localizable features / S. Yun [et al.] // 2019 IEEE/CVF Intern. Conf. on Computer Vision (ICCV), Seoul, Korea (South), 27 Oct. - 2 Nov. 2019. Seoul, 2019. P. 6022–6031. https://doi.org/10.1109/ICCV.2019.00612

5. Cut-thumbnail: A novel data augmentation for convolutional neural network / T. Xie [et al.] // Proc. of the 29th ACM Intern. Conf. on Multimedia, Virtual Event, China, 20–24 Oct. 2021. Virtual Event, China, 2021. P. 1627–1635. https://doi.org/10.1145/3474085.3475302

6. Mixup: Beyond Empirical Risk Minimization / H. Zhang [et al.]. 2018. Mode of access: https://doi.org/10.48550/arXiv.1710.09412. Date of access: 11.08.2024.

7. Person transfer GAN to bridge domain gap for person re-identification / L. Wei [et al.] // 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. Salt Lake City, 2018. P. 79–88. https://doi.org/10.1109/CVPR.2018.00016

8. M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan.arXiv preprint arXiv:1701.07875, 2017.

9. X. Yan, J. Yang, K. Sohn, and H. Lee. Attribute2image:Conditional image generation from visual attributes. arXivpreprint arXiv:1512.00570, 2015.

10. J. Walker, C. Doersch, A. Gupta, and M. Hebert. An uncertain future: Forecasting from static images using variational autoencoders. In European Conference on Computer Vision, pages 835–851. Springer, 2016.

11. ImageNet-Trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness / R. Geirhos [et al.]. 2019. Mode of access: https://doi.org/10.48550/arXiv.1811.12231. Date of access: 11.08.2024.

12. Gong, Y. An Effective Data Augmentation for Person Re-identification / Y. Gong, Z. Zeng. 2021. Mode of access: https://doi.org/10.48550/arXiv.2101.08533. Date of access: 11.08.2024.

13. Adversarially occluded samples for person re-identification / H. Huang [et al.] // 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. Salt Lake City, 2018. P. 5098–5107. https://doi.org/10.1109/CVPR.2018.00535

14. Random Erasing Data Augmentation / Z. Zhong [et al.]. 2020. Mode of access: https://doi.org/ 10.1609/AAAI.V34I07.7000. Date of access: 11.08.2024.

15. DeVries, T. Improved Regularization of Convolutional Neural Networks with CutOut / T. DeVries, G. W. Taylor. 2017. Mode of access: https://doi.org/10.48550/arXiv.1708.04552. Date of access: 11.09.2024.

16. Dropout: A simple way to prevent neural networks from overfitting / N. Srivastava [et al.] // J. of Machine Learning Research. 2014. No. 15. P. 1929–1958. https://doi.org/10.5555/2627435.2670313

17. Choice of activation function in convolution neural network for person re-identification in video surveillance systems / H. Chen [et al.] // Programming and Computer Software. 2022. Vol. 48, № 5. P. 312–321. http://doi.org/10.1134/S0361768822050036

18. Deep residual learning for image recognition / K. He [et al.] // 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. Las Vegas, 2016. P. 770–778. https://doi.org/10.1109/cvpr.2016.90

19. K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In Advances in Neural Information Processing Systems, pages 3483–3491, 2015.

20. M. Arjovsky and L. Bottou. Towards principled methods for training generative adversarial networks. In NIPS 2016 Workshop on Adversarial Training. In review for ICLR, volume 2016, 2017.

Panaskin D.V. APPLICATION OF CVAE AUGMENTATION METHODS TO IMPROVE LEARNING OF NEURAL NETWORKS IN LUNG SOUND ANALYSIS

The article is dedicated to the study of neural networks in the analysis of lung sounds as an additional safeguard for the prevention and development of respiratory diseases, as well as the application of CVAE augmentation methods to improve neural network training. The article evaluates the feasibility of using acoustics, which analyzes noise to detect signs of pulmonary pathologies. It is shown that for the detection of pathologies, it is advisable to use the difference in normalized noise levels and logarithmic scale spectra for the left and right lungs. It is proven that differential noise diagnostics can be based on the use of modified cross-correlation of logarithmic scale spectra, as well as their instantaneous characteristics with the help of neural networks. To preserve the diversity of generated samples, we utilized the advantages of CVAE augmentation methods. To visualize lung sounds, we used an encoder network with real image generation to the latent vector. The generator then provided the raw pixels necessary for reconstruction in accordance with the characteristics of the original images with a given latent vector. Using this augmentation method, it is possible to increase the diversity of training data, which will improve efficiency.

A two-stage neural network training technology (CVAE) is presented, which is distinguished by the use of image augmentation for the preliminary stage and precise adjustment of weight coefficients based on the original image set. In the first stage, training is performed on augmented data, while in the second stage, fine-tuning is carried out, contributing to the improvement of re-identification efficiency. The obtained results can serve as a basis for the creation of a multi-channel automated acoustic noise analysis system for the differential diagnosis of lung conditions, which will allow the development of a modern monitoring system for respiratory diseases in the population.

Key words: *medicine, artificial intelligence, lung diseases, augmentation, diagnostics.*